Case Report

Open Access

CrossMark

Using machine learning for no show prediction in the scheduling of clinical exams

Volume 7 Issue 1 - 2020

Scrivani H, Soeiro F
DASA, Rua Gilberto Sabino 215, Brazil

Correspondence: Hommenig Scrivani, DASA, Rua Gilberto Sabino 215, Sao Paulo, Sao Paulo, Brazil, Tel +55 11 991058062, Email hommenig.scrivani@dasa.com.br

Received: December 13, 2019 | **Published:** February 28, 2020

Abstract

For this work actual data of more than 1.000.000 appointments of large laboratories in São Paulo and Rio de Janeiro (Brazil) were used. The study has the first objective to study the profile of no show patients: patients who schedule a clinical examination but do not attend it. The second objective was to find out which variables are relevant to predict patient non-attendance on the day of examination. As a final objective, it is of interest to provide, at the time of the appointment, whether the patient will attend the scheduled procedure or not. In order to achieve these goals, machine learning based prediction models were developed and analyzed, resulting in a clustering of the scheduled patients according to their no show propensity. Clustering was used instead of a binary result in order to allow for different strategies to be defined as the resulting profiles were analyzed. The models were evaluated by their accuracy and the combination of Random Forest and Logistic Regression techniques presented the best results, identifying clusters which had up to 75% accuracy rate in a validation environment. The main discussions follow in two lines: the first one is related to the allocation of the available time, from a patient who will not attend the examination, to one who will. The second discussion is in the understanding of the needs and circumstances that lead to non-attendance in performing the exam.

Keywords: no show, machine learning, prediction models, clinical exams

Introduction

The present study has the first objective to discover the profile of patients who schedule a clinical examination but do not attend it (who is a no show). The second objective was to find out which variables are relevant to predict patient non-attendance on the day of examination. As a final objective, it is of interest to provide, at the time of the appointment, whether the patient will attend the scheduled procedure or not.

Case presentation

For this work actual data of more than 1,000,000 appointments of large laboratories in São Paulo and Rio de Janeiro (Brazil) were used. Big Data tools and Machine Learning techniques (Random Forest,[1,2] Support Vector Machine[3] and Logistic Regression)[4] were applied to obtain the results. The analyzes were performed using the software R.

The patients were split into clusters based on no show propensity according to the results given by each prediction model, which were studied both individually and in combination with one another. Results were evaluated according to the accuracy level reached in each model and the combination between them worked as a tool for more precise and detailed study and understanding of patient profile clusters.

Discussion

The most relevant results in this experiment were obtained from the combination of Machine Learning techniques, comprising the application of Random Forest models along with Logistic Regression models. The accuracy rate of the Machine Learning model in a validation environment reached 75%, which, combined with the Logistic Regression, allows the definition of different strategies for each group of patients. The main discussions follow in two lines: the first one is related to the allocation of the available time, from a patient who will not attend the examination, to one who will. The second discussion is in the understanding of the needs and circumstances that lead to non-attendance in performing the exam.

Acknowledgments

None.

Conflicts of interest

No financial interest or conflict of interest exists.

References

1. Liaw A. Classification and regression with Random Forest. *randomForest*.
2. Kuhn M. Classification and regression training. *Caret*.
3. Meyer D. Support Vector Machines. *svm*.
4. R-core. Fitting generalized linear models. *glm*.

Citation: Scrivani H, Soeiro F. Using machine learning for no show prediction in the scheduling of clinical exams. *Int J Radiol Radiat Ther.* 2020;7(1):34. DOI: 10.15406/ijrrt.2020.07.00260